

# Statistische und stochastische Grundlagen

VORLESUNGSMITSCHRIEB ZUM MODUL AN DER UNIVERSITÄT STUTTGART

---

# INHALTSVERZEICHNIS

|          |   |           |
|----------|---|-----------|
| <b>I</b> | <b>Statistik</b>  | <b>4</b>  |
| <b>1</b> | <b>Grundbegriffe</b>  | <b>5</b>  |
| 1.1      | Grundbegriffe der Statistik . . . . .                           | 5         |
| 1.2      | Charakterisierung der Merkmale . . . . .                        | 5         |
| 1.3      | Skalen . . . . .  | 5         |
| 1.4      | Datengewinnung, Datenerhebung . . . . .                         | 6         |
| <b>2</b> | <b>Verteilungen und ihre Darstellungen</b>                      | <b>7</b>  |
| 2.1      | Häufigkeiten . . . . .  | 7         |
| 2.2      | Kumulierte Häufigkeiten . . . . .                               | 8         |
| 2.3      | Gruppierung . . . . .   | 8         |
| 2.4      | Lagemaße . . . . .  | 8         |
| 2.4.1    | Arithmetisches Mittel . . . . .                                 | 9         |
| 2.4.2    | Median . . . . .  | 9         |
| 2.4.3    | Modus . . . . .   | 10        |
| 2.4.4    | Geometrisches Mittel . . . . .                                  | 10        |
| 2.4.5    | Harmonisches Mittel . . . . .                                   | 10        |
| 2.5      | Lageregeln . . . . .  | 10        |
| 2.6      | Streuungsmaße . . . . .   | 10        |
| 2.6.1    | Spannweite, Stichprobenspannweite . . . . .                     | 11        |
| 2.6.2    | Mittlere absolute Abweichung vom Median . . . . .               | 11        |
| 2.6.3    | Quantile . . . . .  | 11        |
| 2.6.4    | Fünf-Punkte-Zusammenfassung . . . . .                           | 12        |
| 2.6.5    | Standardabweichung . . . . .                                    | 12        |
| 2.6.6    | Variationskoeffizient . . . . .                                 | 13        |
| 2.6.7    | Varianz, empirische Varianz . . . . .                           | 13        |
| 2.6.8    | Stichprobenvarianz . . . . .                                    | 14        |
| 2.7      | Konzentrationsmaße . . . . .                                    | 14        |
| 2.7.1    | Lorenzkurve . . . . .   | 14        |
| 2.7.2    | Gini-Koeffizient . . . . .                                      | 14        |
| 2.8      | Schiefe und Wölbung . . . . .                                   | 15        |
| <b>3</b> | <b>Multivariate</b>   | <b>16</b> |
| 3.1      | Kontingenztafel . . . . .                                       | 16        |
| 3.2      | Bedingte Häufigkeiten . . . . .                                 | 16        |
| 3.3      | Zusammenhangsmaße . . . . .                                     | 17        |
| 3.3.1    | $\chi^2$ -Koeffizient . . . . .                                 | 17        |
| 3.3.2    | Kontingenzkoeffizient . . . . .                                 | 18        |
| 3.3.3    | Empirischer Korrelationskoeffizient (Bravais-Pearson) . . . . . | 18        |
| 3.3.4    | $\phi$ -Koeffizient . . . . .                                   | 19        |
| 3.4      | Darstellungen von Verteilungen . . . . .                        | 19        |
| 3.4.1    | Lineare Regression . . . . .                                    | 19        |

|           |   |           |
|-----------|---|-----------|
| <b>II</b> | <b>Stochastik</b>   | <b>20</b> |
| <b>1</b>  | <b>Wahrscheinlichkeitsräume</b>                                       | <b>21</b> |
| 1.1       | Allgemeine Definitionen . . . . .                                     | 21        |
| <b>2</b>  | <b>Diskrete Wahrscheinlichkeitsräume</b>                              | <b>22</b> |
| 2.1       | Endliche Wahrscheinlichkeitsräume . . . . .                           | 22        |
| 2.1.1     | Zufallsvariablen auf endlichen Wahrscheinlichkeitsräumen . . . . .    | 23        |
| 2.1.2     | Unabhängigkeit in endlichen Wahrscheinlichkeitsräumen . . . . .       | 25        |
| 2.1.3     | Binomialverteilung . . . . .  | 26        |
| 2.1.4     | Hypergeometrische Verteilung . . . . .                                | 27        |
| 2.2       | Erweiterung auf allgemein diskrete Wahrscheinlichkeitsräume . . . . . | 27        |
| 2.2.1     | Geometrische Verteilung . . . . .                                     | 28        |
| 2.2.2     | Poisson-Verteilung . . . . .  | 29        |
| 2.3       | Grenzwertsätze . . . . .  | 30        |

# 1

Kapitel 1: Statistik

---

# 1: GRUNDBEGRIFFE

## 1.1 Grundbegriffe der Statistik

**Statistische Einheit** Objekte die erfasst werden und an denen die interessierenden Größen erfasst werden

**Grundgesamtheit** Menge aller für die Fragestellung relevanten statistischen Einheiten

**Teilgesamtheit** Teilmenge der Grundgesamtheit

**Stichprobe** Tatsächlich untersuchte Teilmenge der Grundgesamtheit

**Merkmal, Variable** Größe von Interesse

**Merkmalsausprägung, Wert** Konkreter Wert des Merkmals für eine bestimmte statistische Einheit

## 1.2 Charakterisierung der Merkmale

**diskret** Merkmale, die nur endlich viele oder abzählbar unendlich viele Ausprägungen annehmen sind diskret.

**stetig** Merkmale, die Werte aus einem Intervall annehmen können heißen stetig.

**quasi-stetig** Merkmale, die sich nur diskret messen lassen aber aufgrund einer sehr feinen Abstufung wie stetige Merkmale behandelt werden können.

Die Ausprägungen eines stetigen Merkmals lassen sich immer so zusammenfassen, dass es als diskret angesehen werden kann. Die Ausprägungen heißen dann gruppiert oder klassiert.

## 1.3 Skalen

Zusätzlich zur Charakterisierung der Merkmale werden diese anhand ihres Skalenniveaus unterschieden.

**Nominalskala** Wenn die Ausprägungen Namen oder Kategorien sind, die den Einheiten zugeordnet werden heißt das Merkmal *nominalskaliert*. Beispielsweise Geschlecht oder Verwendungszweck.

**Ordinalskala** Merkmale mit Ausprägungen zwar mit Ordnung, bei denen allerdings ein Abstand der Merkmale nicht interpretier- oder vergleichbar ist heißen *ordinalskaliert*. Ein Beispiel hierfür wären Schulnoten.

**Kardinalskala** Ein kardinalskaliertes Merkmal wird oft auch metrisch bezeichnet. Hierbei sind die Abstände der Ausprägungen interpretierbar und zusätzlich ist ein sinnvoller Nullpunkt der Skala festgelegt oder bestimmbar.

Auf Basis dieser Skalenmerkmale nennt man Merkmale mit endlich vielen Ausprägungen, die höchstens ordinalskaliert sind *qualitative* oder *kategoriale Merkmale*. Diese geben eine Qualität aber nicht ein Ausmaß wieder.

Geben die Ausprägungen jedoch eine Intensität oder Ausmaß wieder so spricht man von *quantitativen Merkmalen*. Alle Messungen mit Zahlenwerten stellen Ausprägungen quantitativer Merkmale dar. Ein kardinalskaliertes Merkmal ist stets quantitativ.

## 1.4 Datengewinnung, Datenerhebung

S.18 ff

## 2: VERTEILUNGEN UND IHRE DARSTELLUNGEN

### 2.1 Häufigkeiten

Als *Urliste* bezeichnet man die Menge der Merkmale  $X$  der Untersuchungseinheiten  $U = \{x_1, \dots, x_n\}$ . Die *auf tretenden Ausprägungen* von  $X$  sind die Werte  $\{a_1, \dots, a_k\} \subseteq \{x_1, \dots, x_n\}, k \leq n$ . Oftmals treten in einem großen Datensatz der Größe  $n$  nicht auch  $n$  verschiedene Werte  $x_i$  auf. Damit definieren sich

#### Definition 2.1: Absolute Häufigkeit

Die absolute Häufigkeit einer auftretenden Ausprägung  $a$  in einer Urliste  $U$  ist

$$h(a) = |\{i \in \mathbb{N} \mid x_i = a, x_i \in U\}|.$$

Es gilt immer, dass die Summe aller absoluten Häufigkeiten gleich der Datensatzgröße ist

$$\sum_{i=1}^n h(a_i) = |U|.$$

Die absolute Häufigkeitsverteilung ist dargestellt durch die Folge von Werten

$$h_1, \dots, h_k = h(a_1), \dots, h(a_k)$$

#### Definition 2.2: Relative Häufigkeit

Die relative Häufigkeit einer auftretenden Ausprägung  $a$  in einer Urliste  $U$  ist

$$f(a) = \frac{h(a)}{|U|}.$$

Es gilt ähnlich wie bei der absoluten Häufigkeit für die Summe

$$\sum_{i=1}^n f(a_i) = 1.$$

Eine grafische Darstellung einer Häufigkeitsverteilung nennt man ein *Histogramm*. Bei Histogrammen ist auf die Flächentreue zu achten, das bedeutet, dass der Flächeninhalt der aufgetragenen Rechtecke proportional (oder gleich) zu  $h_j$  oder  $f_j$  ist. So kann das menschliche Auge die Verteilung besser wahrnehmen.

Hat das Histogramm einer Verteilung nur einen deutlich erkennbaren Hochpunkt (Gipfel), heißt sie *unimodal*. Treten mehrere Gipfel auf nennt man die Verteilung *multimodal*. Bei zwei Gipfeln spricht man von einer *bimodalen* Verteilung.

Man nennt eine Verteilung *symmetrisch*, wenn es eine Symmetrieachse gibt, sodass die rechte und linke Hälfte der Verteilung annähernd zueinander spiegelbildlich sind. Eine Verteilung heißt *schief*, wenn

sie deutlich unsymmetrisch ist. Sie heißt dann *linksteil* oder *rechtsschief*, wenn der überwiegende Anteil von Daten linksseitig konzentriert ist. Dann steigt die Verteilung links deutlich steiler ab als rechts. Entsprechend *rechtssteile* oder *linksschiefe* Verteilungen.

## 2.2 Kumulierte Häufigkeiten

Die kumulierten Häufigkeitsverteilungen geben an, wie viele Datenpunkte der Urliste, beziehungsweise welcher Anteil der Daten unterhalb einer Schranke liegen. Um diese Aussage sinnvoll zu beantworten ist zumindest eine Ordinalskala nötig.

### Definition 2.3: Absolute kumulierte Häufigkeitsverteilung

Die absolut kumulierte Häufigkeitsverteilung ist die Funktion

$$H(x) = \sum_{i: a_i \leq x} h_i.$$

### Definition 2.4: Relative kumulierte Häufigkeitsverteilung

Die relative kumulierte Häufigkeitsverteilung oder auch empirische Verteilungsfunktion ist

$$F(x) = \sum_{i: a_i \leq x} f_i.$$

Die kumulierten Häufigkeitsverteilungen sind monoton wachsende, Treppenfunktionen, die an den Sprungstellen rechtsseitig stetig sind.

## 2.3 Gruppierung

Sind alle auftretenden Ausprägungen Elemente eines Intervalls  $[a, b]$ , lässt sich dieses in gleich große Klassen der Größe  $d$  unterteilen.

Eine Klassifizierung ist allgemein

$$[a, c_1), \dots, [c_i, c_{i+1}), [c_{i+1}, c_{i+2}), \dots \quad \forall i : c_{i+1} - c_i = d$$

Klassifizierte Daten sind i.A. einfacher zu interpretieren als große Mengen von Daten, die sich nur wenig voneinander unterscheiden.

Der maximale Fehler bei der Klassifizierung ist die halbe Klassengröße.

## 2.4 Lagemaße

Lagemaße helfen beim Vergleich verschiedener Eigenschaften, bzw dem Vergleich verschiedener statistischer Einheiten mit einer gemeinsamen Eigenschaft.



**Definition 2.5: Lagemaß**

Ein *Lagemaß* ist eine Abbildung  $L : \mathbb{R}^n \rightarrow \mathbb{R}$  mit der Eigenschaft

$$L(x_1 + a, \dots, x_n + a) = L(x_1, \dots, x_n) + a \quad \forall a, x_i \in \mathbb{R} \quad (1 \leq i \leq n)$$

Ein Lagemaß beschreibt das Zentrum einer Verteilung.

Beispiele für Lagemaße sind

**2.4.1 Arithmetisches Mittel**

Das arithmetische Mittel ist nur für quantitative Merkmale sinnvoll. Es berechnet sich durch

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \sum_{i=1}^n (a_i \cdot f_i)$$

aus Rohdaten beziehungsweise aus den Häufigkeitsdaten.

Mit dem arithmetischen Mittel gilt die sogenannte *Schwerpunkteigenschaft*

$$\sum_{i=1}^n (x_i - \bar{x}) = 0.$$

Unter einer linearen Transformation  $x \mapsto ax + b$  verhält sich das arithmetische Mittel analog

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n (ax_i + b) = \frac{a}{n} \sum_{i=1}^n x_i + b = a\bar{x} + b$$

Wie aus der Formel erkennbar, ist das arithmetische Mittel extrem empfindlich gegen Ausreißer. Dafür wurden die folgenden Mittel eingeführt.

**DAS GETRIMMTE MITTEL** Um Ausreißer weniger stark ins Gewicht fallen zu lassen wird der Datensatz absichtlich verkleinert. Beim getrimmten Mittel aus einer sortiert vorliegenden Liste von Daten werden zum Beispiel die oberen und unteren 5% der Daten abgeschnitten, damit fallen auch eventuelle Ausreißer raus. Die Datensatzgröße bleibt jedoch nicht erhalten.

**DAS WINSORISIERTE MITTEL** Ähnlich wie beim getrimmten Mittel wird der Datensatz beim winsorisierten Mittel von oben und unten herein bearbeitet. Anstatt Daten zu löschen werden beispielsweise die oberen 5% durch den nächstkleineren Wert ersetzt. Hierbei bleibt also die Datensatzgröße gleich.

**2.4.2 Median**

Der Median stellt ein robusteres Lagemaß als das arithmetische Mittel dar, er ist resistenter gegen Ausreißer im Datensatz. Für  $x_1 \leq x_2 \leq \dots \leq x_n$ , also einen sortiert vorliegenden Datensatz ist der Median

$$x_{\text{med}} = \begin{cases} x_{\frac{n+1}{2}} & n \text{ ungerade} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & n \text{ gerade} \end{cases}$$

Benötigt eine Zahlenordnung, also eine Ordinalskala.

Der Modus verhält sich unter linearer Transformation  $y = ax + b$  genauso wie das arithmetische Mittel  $y_{\text{med}} = ax_{\text{med}} + b$ .

Mindestens 50% der Daten sind kleiner oder gleich  $x_{\text{med}}$ , genauso sind mindestens 50% der Daten größer oder gleich dem Median  $x_{\text{med}}$ .

### 2.4.3 Modus

Der Modus ist die Ausprägung größter Häufigkeit  $x_{\text{mod}} = a_i$  mit  $h(a_i) = \max \{h(a) \mid a \in A\}$  wobei  $A$  die Menge aller vorkommenden Ausprägungen der Urliste ist. Der Modus ist dann eindeutig, wenn die Häufigkeitsverteilung ein eindeutiges Maximum besitzt.

Der Modus empfiehlt sich schon für nominalskalierte Daten.

Der Modus verhält sich unter linearer Transformation  $y = ax + b$  genauso wie das arithmetische Mittel  $y_{\text{mod}} = ax_{\text{mod}} + b$ .

### 2.4.4 Geometrisches Mittel

Für eine Urliste  $U = \{u_1, \dots, u_n\}$  ist das geometrische Mittel definiert als

$$x_{\text{geom}} = \sqrt[n]{\prod_{i=1}^n u_i}.$$

Das geometrische Mittel wird z.B. bei der Berechnung des effektiven Jahreszinses verwendet, es stellt jedoch kein Lagemaß im engeren Sinne dar.

### 2.4.5 Harmonisches Mittel

Für eine Urliste  $U = \{u_1, \dots, u_n\}$  ist das harmonische Mittel definiert als

$$x_{\text{harm}} = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}}.$$

Genauso wie das geometrische Mittel zählt das harmonische nicht zu den Lagemaßen im engeren Sinne.

## 2.5 Lageregeln

Für symmetrische Verteilungen gilt  $\bar{x} \approx x_{\text{med}} \approx x_{\text{mod}}$ .

Für linkssteile Verteilungen  $\bar{x} > x_{\text{med}} > x_{\text{mod}}$ .

Und ebenso für rechtssteile Verteilungen  $\bar{x} < x_{\text{med}} < x_{\text{mod}}$ .

## 2.6 Streuungsmaße

Um eine Verteilung sinnvoll beschreiben zu können sind zusätzlich zu den Lagemaßen noch Aussagen über die Streuung der Daten um das Mittel nötig.

### Definition 2.6: Streuungsmaß

Ein *Streuungsmaß* ist eine Abbildung  $S : \mathbb{R}^n \rightarrow \mathbb{R}$  für die gilt

$$S(x_1 + a, \dots, x_n + a) = S(x_1, \dots, x_n) \quad \forall a, x_i \in \mathbb{R} (1 \leq i \leq n)$$

Ein Streuungsmaß stellt dar, wie weit gestreut Werte einer Verteilung um ein Mittel liegen.

### 2.6.1 Spannweite, Stichprobenspannweite

Die Stichprobenspannweite stellt dar, in welchem Bereich die Ausprägungen liegen, denn diese ist einfach

$$x_{\max} - x_{\min}.$$

### 2.6.2 Mittlere absolute Abweichung vom Median

Die mittlere absolute Abweichung vom Median stellt eine robustere Alternative zur Stichprobenvarianz dar. Sie ist für eine Urliste  $U = \{x_1, \dots, x_n\}$  definiert als

$$\frac{1}{n} \sum_{i=1}^n |x_i - x_{\text{med}}|$$

### 2.6.3 Quantile

Ein Quantil ist eine Kennzahl, die Daten nach einer relativen Häufigkeit trennt. So trennt das  $p$ -Quantil einer Verteilung die Daten so, dass etwa  $p \cdot 100\%$  der Daten darunter und  $(1 - p) \cdot 100\%$  darüber liegen. Damit ist der Median gerade das 50%-Quantil.

#### Definition 2.7: Quantile

Sei  $U = \{x_1, \dots, x_n\}$  eine geordnete Urliste, d.h.  $x_1 \leq \dots \leq x_j \leq \dots \leq x_n$ . Das  $p$ -Quantil  $x_p$  ist eine Ausprägung  $x_p \in U$  für die gilt

$$\frac{|\{i \in \mathbb{N} \mid x_i \leq x_p, x_i \in U\}|}{n} \geq p \text{ und } \frac{|\{i \in \mathbb{N} \mid x_i \geq x_p, x_i \in U\}|}{n} \geq 1 - p$$

Das heißt es liegen  $p\%$  der Daten unterhalb und  $(1 - p)\%$  der Daten oberhalb des  $p$ -Quantils. Sinnvoll berechnen lässt sich das  $p$ -Quantil durch die Formel

$$x_p = \begin{cases} \frac{1}{2}(x_{(n \cdot p)} + x_{(n \cdot p + 1)}) & \text{falls } n \cdot p \text{ ganzzahlig} \\ x_{(\lfloor n \cdot p \rfloor + 1)} & \text{sonst} \end{cases}$$

Dabei nennt man das 25%-Quantil auch das *untere Quartil* und entsprechend das 75%-Quantil das *obere Quartil*.

#### Definition 2.8: Interquartilsabstand

Für metrische Merkmale ist der sogenannte *Interquartilsabstand* (*interquartile range*) die Distanz

$$d_Q = \text{IQR} = x_{0.75} - x_{0.25}.$$

Der IQR wird zum Beispiel beim Box-Plot verwendet.

Mit dem IQR können Zäune festgelegt werden, außerhalb derer sich höchstwahrscheinlich Ausreißer des Merkmals befinden. Ein Beispiel hierfür ist zum Beispiel der untere Zaun  $z_u = x_{0.25} - 1.5 \cdot d_Q$  und entsprechend die Obergrenze  $z_o = x_{0.75} + 1.5 \cdot d_Q$ , diese Werte werden wiederum beim Box-Plot verwendet.

### 2.6.4 Fünf-Punkte-Zusammenfassung

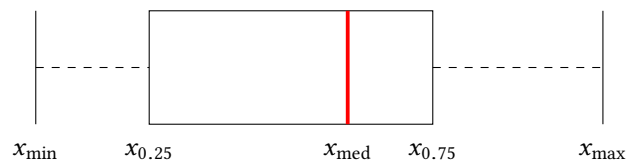
Die Quartile, das Minimum, Maximum sowie der Median teilen den Datensatz in vier Teile, wobei jeder etwa ein Viertel der Merkmale enthält. Die Angabe dieser fünf Werte wird auch als Fünf-Punkte-Zusammenfassung bezeichnet.

#### Box-Plot

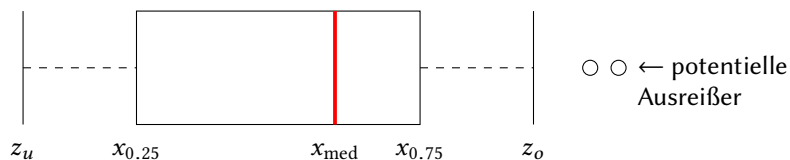
Das Box-Plot ist eine einfache Art und Weise die Ausprägungen einer Verteilung zu visualisieren. Für den Box-Plot wird eine fünf-Punkte-Zusammenfassung,  $(z_u, x_{0.25}, x_{\text{med}}, x_{0.75}, z_o)$  verwendet, wobei  $z_u$  und  $z_o$  beim modifizierten Box-Plot von der klassischen fünf-Punkte-Zusammenfassung abweichen können. Daran werden zwei Definitionen unterschieden, der „normale“ und der modifizierte Box-Plot.

Beim Box-Plot wird ein Rechteck zwischen den Quartilen gezeichnet, das  $x_{\text{med}}$  eingezeichnet als Linie oder Punkt beinhaltet. So sieht man dass sich 50% der Datenpunkte innerhalb der Box befinden. Die nach außen gezeichneten Linien geben an, wie weit die restlichen 50% der Datenpunkte gestreut liegen. Diese sogenannten Whiskers enden in Abhängigkeit von  $z_u$  bzw.  $z_o$ , diese Werte unterscheiden sich bei den beiden Definitionen.

**NORMALER BOX-PLOT** Das Fünftupel besteht aus den Werten  $(x_{\min}, x_{0.25}, x_{\text{med}}, x_{0.75}, x_{\max})$ . Wobei  $x_{\min}$  und  $x_{\max}$  die kleinste und größte Ausprägung der Verteilung darstellen. So sind alle Werte in der Spannweite der Whiskers enthalten. Ein Box-Plot sieht dann wie folgt aus



**MODIFIZIERTER BOX-PLOT (NACH FAHRMEIR)** Der wichtigste Unterschied zum normalen Box-Plot ist dass, anstatt der minimalen und maximalen Werte für  $z_u, z_o$  ein Zaun gewählt wurde. So ist  $z_u = x_{0.25} - 1.5 \cdot d_Q$  und  $z_o = x_{0.75} + 1.5 \cdot d_Q$  wobei  $d_Q$  der Interquartilsabstand ist. Allerdings ist zu beachten dass die Whiskers von der größten/ kleinsten Ausprägung innerhalb des Zauns zur Box ausgehen. Liegen also beispielweise innerhalb des Bereichs  $[z_u, x_{0.25}]$  keine Datenwerte, so existiert kein unterer Whisker. Datenpunkte, die außerhalb des Zauns liegen, werden mit Punkten dargestellt. Dies ist ein gutes Anzeichen für eventuelle Ausreißer.



### 2.6.5 Standardabweichung

Die Standardabweichung ist für eine Urliste  $U = \{x_1, \dots, x_n\}$  von metrischen Daten definiert als

$$\tilde{s} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\sum_{i=1}^n (a_i - \bar{x})^2 \cdot f_i}$$

wobei die  $a_i$  die Ausprägungen der Urliste sind und  $f_i$  die relative Häufigkeit der Ausprägung  $a_i$  ist. Unter einer linearen Transformation  $x \mapsto ax + b$  der Ausprägungen wird  $\tilde{s}$  um  $|a|$  gedehnt. Eine Verschiebung der Werte um  $b$  hat keine Auswirkung.

### 2.6.6 Variationskoeffizient

Der Variationskoeffizient ist eine maßstabsunabhängige Maßzahl für die Streuung, sie basiert auf der Standardabweichung und ist definiert als

$$v = \frac{\tilde{s}}{\bar{x}}, \quad \bar{x} > 0$$

### 2.6.7 Varianz, empirische Varianz

Die empirische Varianz ist das Quadrat der Standardabweichung  $\tilde{s}^2$ .

$$\tilde{s}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Sowohl Standardabweichung als auch die Varianz sind nicht resistent, reagieren also sehr empfindlich auf Ausreißer.

Für eine Berechnung von Hand gilt der sogenannte Verschiebungssatz

#### Satz 2.9: Verschiebungssatz

Für jedes  $c \in \mathbb{R}$  gilt

$$\sum_{i=1}^n (x_i - c)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - c)^2.$$

Damit gilt insbesondere mit  $c = 0$  für die Varianz

$$\tilde{s}^2 = \left( \frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \bar{x}.$$

**BEWEIS:**

$$\begin{aligned} \sum_{i=1}^n (x_i - c)^2 &= \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - c)^2 \\ &= \sum_{i=1}^n \left[ (x_i - \bar{x})^2 + 2(x_i - \bar{x})(\bar{x} - c) + (\bar{x} - c)^2 \right] \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + \underbrace{2(\bar{x} - c) \sum_{i=1}^n (x_i - \bar{x})}_{=0} + \sum_{i=1}^n (\bar{x} - c)^2 \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - c)^2 \end{aligned}$$

**LINEARE TRANSFORMATION** Mit einer linearen Abbildung der Ausprägungen  $y_i = ax_i + b$  verhält sich die Varianz der Daten  $y_i$

$$\tilde{s}_y^2 = a^2 \tilde{s}_x^2 \text{ bzw. } \tilde{s}_y = |a| \tilde{s}_x$$

dies ergibt sich direkt aus den Formeln für die Varianz

$$\begin{aligned}\tilde{s}_y^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \\ &= \frac{1}{n} \sum_{i=1}^n (ax_i + b - a\bar{x} - b)^2 \\ &= a^2 \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = a^2 \tilde{s}_x^2\end{aligned}$$

### 2.6.8 Stichprobenvarianz

Die Stichprobenvarianz stellt ein nicht resistentes Streuungsmaß dar. Sie ist für eine Urliste  $U = \{x_1, \dots, x_n\}$  definiert als

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2$$

## 2.7 Konzentrationsmaße

Konzentration geben Aufschluss über die Stärke der Konzentration von Daten.

### 2.7.1 Lorenzkurve

Ausgehend von geordneten Ausprägungen  $x_1 \leq \dots \leq x_n$  stellt die Lorenzkurve den Anteil der kumulierten relativen Merkmalssumme bezüglich dem Anteil der Merkmalsträger von der Grundgesamtheit dar.

#### Definition 2.10: Lorenzkurve

Die Lorenzkurve ergibt sich als Streckenzug durch die Punkte  $(0, 0), (u_1, v_1), \dots, (1, 1)$  wobei

$$u_j = \frac{j}{n} \quad \text{und} \quad v_j = \frac{\sum_{i=1}^j x_i}{\sum_{i=1}^n x_i}$$

ist.

Der Grundgedanke ist, darzustellen auf welchen Teil der Merkmalsträger welcher Anteil der Merkmalssumme zurückgeht.

Die Lorenzkurve wächst immer monoton und konvex, d.h. sie wölbt sich nach unten.

### 2.7.2 Gini-Koeffizient

In der Lorenzkurve drückt sich Konzentration der Daten durch Entfernung von der ersten Winkelhalbierenden aus. Genau dies nutzt der Gini-Koeffizient  $G$  aus.

$G$  ist gleich dem Verhältnis zwischen dem von Lorenzkurve und Diagonale eingeschlossenen Flächeninhalt und der Fläche unter der Winkelhalbierenden.

$$G = \frac{2 \sum_{i=1}^n i x_i}{n \sum_{i=1}^n x_i} - \frac{n+1}{n} \quad G \in [0, \infty)$$

Dabei ist der minimale Wert des Gini-Koeffizienten  $G_{\min} = 0$  und das Maximum ist  $G_{\max} = \frac{n-1}{n}$ . Da der Wert des Gini-Koeffizienten mit von der Eingabegröße abhängen kann, bietet sich der normierte Gini-Koeffizient für Vergleiche an

$$G^* = \frac{G}{G_{\max}} = \frac{n}{n-1}G \quad G^* \in [0, 1]$$

## 2.8 Schiefe und Wölbung

## 3: MULTIVARIATE

Bisher wurden nur eindimensionale Daten erfasst, nicht aber verschiedene Merkmale in Zusammenhang gebracht und gemeinsam betrachtet oder miteinander verglichen.

### 3.1 Kontingenztafel

Die Kontingenztafel eignet sich zur Darstellung der gemeinsamen Verteilung von zwei diskreten Merkmalen mit relativ wenigen Ausprägungen.

Auf Basis der Ausprägungen  $a_1, \dots, a_k$  des Merkmals  $X$  und  $b_1, \dots, b_m$  für  $Y$  liegen in der Urliste die gemeinsamen Messwerte vor. Das heißt die Urliste besteht aus den Tupeln  $(a_i, b_j)$ . Analog zum Eindimensionalen sind die absoluten Häufigkeiten  $h_{ij}$  definiert. Darauf aufbauend ebenfalls völlig analog die relativen Häufigkeiten  $f_{ij}$ .

**KONTINGENZTAFEL DER ABSOLUTEN HÄUFIGKEITEN** Die aus diesen Werten entstehende Tafel heißt  $(k \times m)$ -Kontingenztafel der absoluten Häufigkeiten. Sie enthält neben den Häufigkeitsdaten zusätzlich noch die Spalten- beziehungsweise Zeilensummen der Werte.

|          | $b_1$         | $\dots$ | $b_m$         |                                    |
|----------|---------------|---------|---------------|------------------------------------|
| $a_1$    | $h_{11}$      | $\dots$ | $h_{1m}$      | $h_{1\cdot} = \sum_{i=1}^m h_{1i}$ |
| $a_2$    | $h_{21}$      | $\dots$ | $h_{2m}$      | $h_{2\cdot} = \sum_{i=1}^m h_{2i}$ |
| $\vdots$ | $\vdots$      | $\dots$ | $\vdots$      |                                    |
| $a_k$    | $h_{k1}$      | $\dots$ | $h_{km}$      | $h_{k\cdot} = \sum_{i=1}^m h_{ki}$ |
|          | $h_{\cdot 1}$ |         | $h_{\cdot m}$ | $n$                                |

Die Zeilensummen  $h_{i\cdot}$  werden auch als Randhäufigkeiten des Merkmals  $X$  bezeichnet. Diese Werte sind die einfachen Häufigkeiten mit denen das Merkmal  $X$  die Werte  $a_1, \dots, a_k$  annimmt, wenn  $Y$  nicht berücksichtigt wird.

Analog dazu sind die Spaltensummen die Häufigkeiten von  $Y$  unter Vernachlässigung des Merkmals  $X$ .

**KONTINGENZTAFEL DER RELATIVEN HÄUFIGKEITEN** Da Anteile beziehungsweise Prozente häufig anschaulicher sind als absolute Häufigkeitswerte betrachtet man häufig auch die Häufigkeitstafel der relativen Häufigkeiten. Diese entsteht durch teilen durch die Gesamtzahl  $n$ .

|          | $b_1$         | $\dots$ | $b_m$         |                                    |
|----------|---------------|---------|---------------|------------------------------------|
| $a_1$    | $f_{11}$      | $\dots$ | $f_{1m}$      | $f_{1\cdot} = \sum_{i=1}^m f_{1i}$ |
| $a_2$    | $f_{21}$      | $\dots$ | $f_{2m}$      | $f_{2\cdot} = \sum_{i=1}^m f_{2i}$ |
| $\vdots$ | $\vdots$      | $\dots$ | $\vdots$      |                                    |
| $a_k$    | $f_{k1}$      | $\dots$ | $f_{km}$      | $f_{k\cdot} = \sum_{i=1}^m f_{ki}$ |
|          | $f_{\cdot 1}$ |         | $f_{\cdot m}$ | 1                                  |

### 3.2 Bedingte Häufigkeiten

Aus den gemeinsamen Häufigkeiten lässt sich nicht direkt auf den Zusammenhang zweier Merkmale schließen. So kann man ein Merkmal fest wählen und dann die Häufigkeitsverteilung des anderen



Merkmals unabhängig davon betrachten, mit dieser Herangehensweise kommt man zu den bedingten Häufigkeiten.

**Definition 3.1: Bedingte Häufigkeiten**

Für zwei Merkmale  $X$  und  $Y$  mit den Ausprägungen  $a_1, \dots, a_k$  und  $b_1, \dots, b_m$  ist

$$f_Y(b_j|a_i) = \frac{h_{ij}}{h_i}$$

die Häufigkeit des Merkmals  $b_j$  aus  $Y$  unter der Bedingung  $X = a_i$ .

Daraus geht durch die Werte

$$f_Y(b_1|a_i), \dots, f_Y(b_m|a_i)$$

die bedingte Häufigkeitsverteilung von  $Y$  unter der Bedingung  $X = a_i$  (Kurzschreibweise:  $(Y|X = a_i)$ ) hervor.

Analog für eine fest gewählte Ausprägung  $Y = b_i$  die bedingte Häufigkeitsverteilung  $f_X(a_j|b_i)$ .

### 3.3 Zusammenhangsmaße

Wir betrachten zunächst wie sich zwei Merkmale zueinander verhalten würden, wenn keinerlei Zusammenhang zwischen ihnen bestünde.

In einer Kontingenztafel müssten sich dann die einzelnen Spalten proportional zu den Spaltensummen verhalten und analog die Zeilen zu den Zeilensummen. Daraus ergibt sich die *erwartete Häufigkeit einer Ausprägung bei unabhängigen Merkmalen*  $X$  und  $Y$

$$\tilde{h}_{ij} = \frac{h_{i.} \cdot h_{.j}}{n}$$

Um den Zusammenhang zweier Merkmale zu untersuchen betrachten wir also den Unterschied zwischen den tatsächlichen Häufigkeiten  $h_{ij}$  und den jeweils erwarteten Häufigkeiten  $\tilde{h}_{ij}$ .

#### 3.3.1 $\chi^2$ -Koeffizient

Basierend auf der oben angesprochenen Differenz wird das erste Zusammenhangsmaß konstruiert.

**Definition 3.2:  $\chi^2$ -Koeffizient**

Für zwei Merkmale  $X$  und  $Y$  mit den Ausprägungen  $a_1, \dots, a_k$  und  $b_1, \dots, b_m$  ist

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(h_{ij} - \tilde{h}_{ij})^2}{\tilde{h}_{ij}} \quad \chi^2 \in [0, \infty)$$

Ist  $\chi^2$  groß, weichen die Häufigkeiten also stark von der Erwartung ab, hängen die Merkmale vermutlich voneinander ab. Sind die Abweichungen allerdings relativ klein und damit  $\chi^2$  ebenfalls, so sind die Merkmale wahrscheinlich unabhängig. Selbst bei unabhängigen Merkmalen ist meist wegen zufälligem Rauschen  $\chi^2 \neq 0$ , eine Entscheidung ist so also nicht möglich.

### 3.3.2 Kontingenzkoeffizient

Problematisch am  $\chi^2$ -Koeffizienten ist die Abhängigkeit von der Dimension der Tafel. Es kann nicht ohne weiteres aus dem Wert des Koeffizienten auf eine Unabhängigkeit der Merkmale geschlossen werden. Als erster Normierungsschritt folgt daraus der Kontingenzkoeffizient.

#### Definition 3.3: Kontingenzkoeffizient

Aus dem  $\chi^2$ -Koeffizienten für zwei Merkmale mit der Urliste der Größe  $n$  ist der Kontingenzkoeffizient

$$K = \sqrt{\frac{\chi^2}{n + \chi^2}} \quad K \in [0, K_{\max}] \text{ für } K_{\max} = \sqrt{\frac{M-1}{M}}, \quad M = \max\{m, k\}$$

Dabei seien  $m$  und  $k$  die Mächtigkeit der Ausprägungsliste der beiden Merkmale.

**KORRIGIERTER KONTINGENZKOEFFIZIENT** Da auch dieser für einen sinnvollen Vergleich nicht ausreichend normiert ist, wird der korrigierte Kontingenzkoeffizient eingeführt, wobei  $K$  durch  $K_{\max}$  normiert wird.

$$K^* = \frac{K}{K_{\max}} \quad K^* \in [0, 1]$$

Sowohl der Kontingenzkoeffizient als auch der  $\chi^2$ -Koeffizient stellen nur die Stärke des Zusammenhangs dar, nicht aber eine Richtung der Wirkungsweise.

### 3.3.3 Empirischer Korrelationskoeffizient (Bravais-Pearson)

Um nicht nur die Stärke des Zusammenhangs zu untersuchen sondern auch die Richtung der Wirkungsweise, wird der empirische Korrelationskoeffizient auf metrischen Merkmalen eingeführt.

#### Definition 3.4: Emp. Korrelationskoeffizient nach Bravais-Pearson

Für zwei metrische Merkmale  $X$  und  $Y$  auf einer Urliste der Größe  $n$  ist der empirische Korrelationskoeffizient

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\tilde{s}_{XY}}{\tilde{s}_X \cdot \tilde{s}_Y}$$

wobei  $\tilde{s}_X$  bzw.  $\tilde{s}_Y$  die Standardabweichungen der Merkmale  $X$  und  $Y$  sind.

Der Korrelationskoeffizient misst die Stärke des linearen Zusammenhangs. Damit ergibt sich für  $r > 0$  eine positive Korrelation also ein gleichsinniger Zusammenhang und ein gegensinniger Zusammenhang für  $r < 0$ . Bei  $r \approx 0$  sind die Merkmale nicht korreliert.

**EMPIRISCHE KOVARIANZ** Die Kovarianz ist die Summe der Abweichungsprodukte mit  $n$  normiert.

$$\tilde{s}_{XY} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

**RECHENGÜNSTIGE VARIANTE DES KORRELATIONSKOEFFIZIENTEN**

$$r = \frac{\sum_{i=1}^n (x_i y_i) - n \bar{x} \bar{y}}{\sqrt{\left( \sum_{i=1}^n x_i^2 - n \bar{x}^2 \right) \cdot \left( \sum_{i=1}^n y_i^2 - n \bar{y}^2 \right)}}$$

**3.3.4  $\phi$ -Koeffizient**

Für Merkmale mit nur zwei Ausprägungen (*dichotome* oder *binäre* Merkmale) wird der Bravis-Pearson-Koeffizient auch  $\phi$ -Koeffizient genannt. Er berechnet sich aus den Häufigkeitseinträgen der Kontingenztafel

$$r = \frac{h_{11} \cdot h_{22} - h_{12} \cdot h_{21}}{\sqrt{h_{1.} h_{2.} h_{.1} h_{.2}}} = \phi$$

Es gilt sogar zusätzlich

$$\phi^2 = \frac{\chi^2}{n}$$

**3.4 Darstellungen von Verteilungen****3.4.1 Lineare Regression**

# 2

Kapitel 2: Stochastik

---

# 1: WAHRSCHEINLICHKEITSRÄUME

## 1.1 Allgemeine Definitionen

- Ein Wahrscheinlichkeitsraum besteht aus einem *Grundraum*  $\Omega$ , einem System  $\mathcal{A}$  von Teilmengen von  $\Omega$  und der *Wahrscheinlichkeitsverteilung*  $P$ .
- Der Grundraum  $\Omega \neq \emptyset$  ist nichtleer und ist die Menge der Ergebnisse. Elemente  $A \in \mathcal{A}$  heißen Ereignisse, ein Ereignis mit  $|A| = 1$  heißt Elementarereignis. Außerdem ist  $\emptyset \in \mathcal{A}$  das *unmögliche Ereignis* und  $\Omega \in \mathcal{A}$  das sogenannte *sichere Ereignis*.
- Die Wahrscheinlichkeitsverteilung ist eine Abbildung  $P : \mathcal{A} \rightarrow \mathbb{R}$ .
- Die Ereignismenge  $\mathcal{A}$  muss unter den Mengenoperationen  $\{\cap, \cup, \setminus\}$  und Komplement abgeschlossen sein. Man nennt zwei Ereignisse  $A, B \in \mathcal{A}$  *unvereinbar*, wenn  $A \cap B = \emptyset$  ( $A$  und  $B$  sind disjunkte Mengen). Für paarweise unvereinbare Ereignisse  $A_i$  definieren wir  $\bigcup_i A_i := \sum_i A_i$ .
- Eine *Zufallsvariable* ist eine Abbildung  $X : \Omega \rightarrow \mathbb{R}$ , sie ordnet einem Ergebnis einen Wert zu. Zum Beispiel wäre die Augensumme beim Werfen mehrerer Würfel eine Zufallsvariable. Für eine Zufallsvariable  $X$  sei  $X^{-1}$  für eine beliebige Teilmenge  $A \in \mathcal{A}$  definiert als das Urbild  $X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}$ .

**KURZSCHREIBWEISEN:** Für das Urbild eines Ereignisses bezüglich einer Zufallsvariable wird eine Kurzschreibweise definiert. So ist zum Beispiel  $\{3 \leq X \leq 4\} := X^{-1}([3, 4])$ . Auch eine Schreibweise wie  $\{X = 3\} \cup \{X = 4\} := X^{-1}(\{3, 4\})$  ist möglich.

## 2: DISKRETE WAHRSCHEINLICHKEITSRÄUME

Einen einfachen Einstieg in die Wahrscheinlichkeitsräume stellen die endlichen Wahrscheinlichkeitsräume dar. Sie sind ein Sonderfall der diskreten Wahrscheinlichkeitsräume.

### 2.1 Endliche Wahrscheinlichkeitsräume

Ein Wahrscheinlichkeitsraum ist endlich, wenn  $|\Omega| < \infty$  ist. Man verwendet dann üblicherweise  $\mathcal{A} = \text{Pot}(\Omega)$ . Damit ist  $\mathcal{A}$  unter allen Mengenoperationen abgeschlossen und für eine Zufallsvariable  $X$  ist  $X^{-1}(B) \in \mathcal{A}$  für alle  $B \subseteq \mathbb{R}$ .

#### Definition 2.1: Endlicher Wahrscheinlichkeitsraum

Das Paar  $(\Omega, P)$  ist ein endlicher Wahrscheinlichkeitsraum, wenn gilt

**EW1**  $\forall A \subseteq \Omega : P(A) \geq 0$

**EW2**  $P(\Omega) = 1$

**EW3**  $\forall A, B \subseteq \Omega, A \cap B = \emptyset : P(A \cup B) := P(A + B) = P(A) + P(B)$

#### Satz 2.2: Folgerungen aus der Definition

1.  $P(\emptyset) = 0$
2.  $P(\sum_i A_i) = \sum_i P(A_i)$
3.  $0 \leq P(A) \leq 1$
4.  $\forall A \subseteq \Omega : P(\bar{A}) = 1 - P(A)$
5.  $A \subseteq B \subseteq \Omega \Rightarrow P(A) \leq P(B)$
6.  $\forall A, B \subseteq \Omega : P(A \cup B) = P(A) + P(B) - P(A \cap B)$
7.  $P(\bigcup_i A_i) \leq \sum_i P(A_i)$
8.  $P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$

**BEISPIEL:** Wir werfen einen Oktaeder, das heißt  $\Omega = \{1, \dots, 8\}$  und  $\mathcal{A} = \text{Pot } \Omega$ . Es gilt  $p(\omega) = P(\{\omega\}) = \frac{1}{8}$  für alle  $\omega \in \Omega$ . Wir definieren das Ereignis  $A_1 = \{2, 3, 5, 7\}$ , also alle Primzahlen auf dem Würfel.

$$P(A_1) = p(2) + p(3) + p(5) + p(7) = \sum_{\omega \in A_1} p(\omega) = \frac{1}{2}.$$

Analog ist für  $A_2 = \{1, 3, 5, 7\}$  die Wahrscheinlichkeit  $P(A_2) = \frac{1}{2}$ . Für die Vereinigung gilt

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = \frac{1}{2} + \frac{1}{2} - \frac{3}{8} = \frac{5}{8}.$$

### 2.1.1 Zufallsvariablen auf endlichen Wahrscheinlichkeitsräumen

Zufallsvariablen auf endlichen Wahrscheinlichkeitsräumen sind dann Abbildungen  $X : \Omega \rightarrow \mathbb{R}$  mit der Verteilung  $P^X : \text{Pot}(\Omega) \rightarrow \mathbb{R}$  wobei  $P^X(B) := P(X^{-1}(B))$  für alle  $B \subseteq X(\Omega)$  ist.

#### KURZSCHREIBWEISEN:

- Völlig analog zu den bereits eingeführten Kurzschreibweisen definieren wir

$$\begin{aligned} P^X(B) &= P(X^{-1}(B)) = P(\{\omega \in \Omega \mid X(\omega) \in B\}) \\ &= P(\{X \in B\}) \\ &= P(X \in B) \end{aligned}$$

- Darauf aufbauend ist  $P(a \leq X \leq b)$  für  $a, b \in \mathbb{R}$  die Wahrscheinlichkeit  $P^X([a, b] \cap X(\Omega))$ .
- Und für einelementige Mengen ist  $P(X = x) = P(X \in \{x\})$ .

#### Definition 2.3: Verteilungsfunktion

Eine Verteilungsfunktion der Zufallsvariablen  $X$  ist eine Abbildung  $F^X : \mathbb{R} \rightarrow \mathbb{R}$  mit  $F^X(x) = P(X \leq x)$ . Dies ist eine monoton steigende, rechtsseitig stetige Funktion.

#### Definition 2.4: Laplace-Experiment

Ein Laplace-Experiment entspricht der intuitiven Gleichverteilung. Es handelt sich um einen Wahrscheinlichkeitsraum  $(\Omega, P)$  wobei  $p(\omega) = \frac{1}{|\Omega|}$  für alle  $\omega \in \Omega$  gilt.

Das bedeutet  $\forall A \subseteq \Omega : P(A) = \frac{|A|}{|\Omega|}$ .

**BEMERKUNG:** Für eine Zufallsvariable  $X$  ist  $P^X$  in der Regel keine Gleichverteilung, da  $|X^{-1}(\{\omega\})| > 1$  vorkommen kann.

**BEISPIEL:** Für zwei Würfel ist der Grundraum  $\Omega = \{1, \dots, 6\}^2$  und die Wahrscheinlichkeit der Elementarereignisse ist  $p((\omega_1, \omega_2)) = \frac{1}{36}$  für alle  $(\omega_1, \omega_2) \in \Omega$ . Die Zufallsvariable  $M(\omega_1, \omega_2) = \max\{\omega_1, \omega_2\}$  beschreibt das Maximum der beiden Würfelresultate. Damit ist

|            |                |                |                |                |                |                 |
|------------|----------------|----------------|----------------|----------------|----------------|-----------------|
| $m$        | 1              | 2              | 3              | 4              | 5              | 6               |
| $P(M = m)$ | $\frac{1}{36}$ | $\frac{3}{36}$ | $\frac{5}{36}$ | $\frac{7}{36}$ | $\frac{9}{36}$ | $\frac{11}{36}$ |

#### Definition 2.5: Erwartungswert einer Zufallsvariablen

Der Erwartungswert einer Zufallsvariablen  $X$  auf dem Wahrscheinlichkeitsraum  $(\Omega, P)$  ist

$$\begin{aligned} E(X) &= \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\}) \\ &= \sum_{x \in X(\Omega)} x \cdot P(X = x) \end{aligned}$$

**Satz 2.6: Sätze zum Erwartungswert**

Für zwei Zufallsvariablen  $X$  und  $Y$  auf dem Wahrscheinlichkeitsraum  $(\Omega, P)$  und  $a \in \mathbb{R}$  gilt

**Linearität 1**  $E(X + Y) = E(X) + E(Y)$

**Linearität 2**  $E(a \cdot X) = a \cdot E(X)$

**Indikatorfunktion**  $E(1_A) = P(A)$  mit  $1_A = 1 \Leftrightarrow X(\omega) \in A$

**Ordnung**  $\forall \omega \in \Omega : X(\omega) \leq Y(\omega) \Rightarrow E(X) \leq E(Y)$

**Definition 2.7: Varianz und Standardabw. von Zufallsvariablen**

Die Varianz einer Zufallsvariablen  $X$  auf dem Wahrscheinlichkeitsraum  $(\Omega, P)$  ist

$$V(X) = E((X - E(X))^2).$$

Entsprechend zur Standardabweichung von Merkmalen aus der Statistik ist die Standardabweichung einer Zufallsvariablen die Wurzel der Varianz

$$\sigma(X) = \sqrt{V(X)}.$$

**Satz 2.8: Sätze zur Varianz**

Für eine Zufallsvariable  $X$  und alle  $\alpha, \beta \in \mathbb{R}$  gilt für die Varianz

1.  $V(\alpha X + \beta) = \alpha^2 V(X)$  (Lineare Transformation)
2.  $V(X) = E(X^2) - (E(X))^2$  (Verschiebungssatz)
3.  $V(X) \geq 0$
4.  $V(X) = 0 \Leftrightarrow \exists! a \in \mathbb{R} : P(X = a) = 1$

**BEWEIS:**

1. Varianz einer linear transformierten Zufallsvariable

$$\begin{aligned} V(\alpha X + \beta) &= E[((\alpha X + \beta) - E(\alpha X + \beta))^2] \\ &= E[(\alpha X - \alpha E(X))^2] && \text{(Linearität)} \\ &= E[\alpha^2 (X - E(X))^2] \\ &= \alpha^2 E[(X - E(X))^2] \\ V(\alpha X + \beta) &= \alpha^2 V(X) \end{aligned}$$

2. Beweis des sog. Verschiebungssatzes mit der Linearität des Erwartungswerts

$$\begin{aligned} V(X) &= E[(X - E(X))^2] \\ &= E[X^2 - 2X \cdot E(X) + E(X)^2] \\ &= E(X^2) - 2E(X) \cdot E(E(X)) + E(X)^2 \\ &= E(X^2) - E(X)^2 \end{aligned}$$



**Satz 2.9: Tschebyscheff-Ungleichung**

Für eine Zufallsvariable  $X$  auf dem Wahrscheinlichkeitsraum  $(\Omega, P)$  gilt für alle  $\epsilon > 0$

$$P(|X - E(X)| \geq \epsilon) \leq \frac{V(X)}{\epsilon^2}$$

**Definition 2.10: Bedingte Wahrscheinlichkeit**

Auf Basis eines Zufallsexperiments wird die Wahrscheinlichkeit, dass ein Ereignis  $B$  eintritt, während zusätzlich ein fest gewähltes Ereignis  $A$  eingetreten ist, *bedingte Wahrscheinlichkeit* bezeichnet. Man interessiert sich also für die Wahrscheinlichkeit des Ereignisses  $B$  unter der Bedingung, dass  $A$  bereits eingetreten ist. In Zeichen ist die gesuchte Wahrscheinlichkeit

$$P(B|A) = P_A(B) := \frac{P(B \cap A)}{P(A)}$$

**2.1.2 Unabhängigkeit in endlichen Wahrscheinlichkeitsräumen****Definition 2.11: Unabhängigkeit von Ereignissen**

Man nennt zwei Ereignisse  $A, B \subset \Omega$  *unabhängig*, wenn

$$P(A \cap B) = P(A) \cdot P(B)$$

gilt.

**BEMERKUNG:**

- Die Unabhängigkeit von Ereignissen ist offensichtlich eine symmetrische Relation.
- Unabhängigkeit ist nicht mit stochastischer Kausalität zu verwechseln
- Sind zwei Ereignisse unabhängig, dann ist die bedingte Wahrscheinlichkeit gleich der totalen
 
$$P(A|B) = P(A).$$
- Sind zwei Ereignisse  $A, B$  unvereinbar, das heißt  $A \cap B = \emptyset$ , sind sie genau dann unabhängig, wenn  $P(A) = 0$  oder  $P(B) = 0$  gilt.
- Sind zwei Ereignisse  $A, B$  unabhängig, so sind auch deren Komplemente  $\bar{A}, \bar{B}$  unabhängig.

**Definition 2.12: Unabhängigkeit von Zufallsvariablen**

Zwei Zufallsvariablen in einem Wahrscheinlichkeitsraum  $(\Omega, P)$  heißen *unabhängig*, wenn für alle  $B_X, B_Y \subseteq \Omega$  gilt

$$\begin{aligned} P(X \in B_X \wedge Y \in B_Y) &= P(X \in B_X) \cdot P(Y \in B_Y) \\ P(X^{-1}(B_X) \cap Y^{-1}(B_Y)) &= P(X^{-1}(B_X)) \cdot P(Y^{-1}(B_Y)) \end{aligned}$$

**BEMERKUNG:** Es reicht zwar diese Eigenschaft nur für die einelementigen Teilmengen von  $\Omega$  zu zeigen, allerdings entstehen auch damit bereits zu viele Kombinationen.

### Satz 2.13: Erwartungswert bei unabhängigen ZV

Für zwei unabhängige Zufallsvariablen gilt für das Produkt der beiden

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

**BEWEIS:** Für das Produkt zweier unabhängiger Zufallsvariablen ist der Erwartungswert

$$\begin{aligned} E(X \cdot Y) &= \sum_{z \in (X \cdot Y)(\Omega)} z \cdot P(X \cdot Y = z) \\ &= \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} xy \cdot P(X = x \wedge Y = y) \\ &= \sum_{x \in X(\Omega)} \left[ x \cdot P(X = x) \cdot \sum_{y \in Y(\Omega)} y \cdot P(Y = y) \right] && \text{(Unabhängigkeit)} \\ &= \sum_{x \in X(\Omega)} x \cdot P(X = x) \cdot \sum_{y \in Y(\Omega)} y \cdot P(Y = y) \\ &= E(X) \cdot E(Y) \end{aligned}$$

## 2.1.3 Binomialverteilung

### Definition 2.14: Bernoulli-Experiment

Ein Bernoulli-Experiment ist ein Zufallsexperiment mit genau zwei möglichen Ausgängen. Das heißt für eine Zufallsvariable  $X$  gilt  $X(\Omega) = \{0, 1\}$  mit den Wahrscheinlichkeiten

$$P(X = 1) = p \qquad P(X = 0) = (1 - p) = q$$

**BEMERKUNG:** Der Erwartungswert eines Bernoulli-Experiments ist  $E(X) = p$ . Für die Varianz gilt

$$V(X) = E(X)^2 - 2p \cdot E(X) + p^2 = p - p^2 = p(1 - p) = pq$$

Führt man nun  $n$  *unabhängige* Bernoulli-Experimente  $X_i, i \in \{1, \dots, n\}$  durch, erhält man eine sogenannte *Bernoulli-Kette* der Länge  $n$ . Ein Ereignis dieses Grundraums ist  $\omega = (\omega_1, \dots, \omega_n)$  wobei  $X_i(\omega) = \omega_i$  gilt. Der Grundraum der Bernoulli-Kette ist also  $\Omega = \{0, 1\}^n$ . Damit gilt für die Verteilung des Wahrscheinlichkeitsraums

$$\begin{aligned} P(\{(\omega_1, \dots, \omega_n)\}) &= P(X_1 = \omega_1, X_2 = \omega_2, \dots, X_n = \omega_n) \\ &= \prod_{i=1}^n P(X_i = \omega_i) && \text{(Unabhängigkeit)} \\ &= \prod_{i: \omega_i=1} p \cdot \prod_{i: \omega_i=0} (1-p) \end{aligned}$$

Damit ist bei  $n$  Durchführungen die Wahrscheinlichkeit  $k$  mal das Ergebnis  $\omega_i = 1$  zu erhalten ( $k$  Treffer bei  $n$  Versuchen)

$$p^k \cdot (1-p)^{n-k}$$

Betrachten wir nun also die Zufallsvariable  $X$ , die die Summe der Treffer beschreibt  $X : \sum_{i=1}^n X_i$ . Das Ereignis  $\{X = k\} = \{(\omega_1, \dots, \omega_n) \in \Omega \mid \sum_{i=1}^n \omega_i = k\}$  beschreibt die Ausgänge, bei denen genau  $k$  von  $n$  Treffern aufgetreten sind. Es ist  $|\{X = k\}| = \binom{n}{k}$ . Damit erhält man insgesamt

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}.$$

Man nennt  $X$  dann *binomialverteilt* mit den Parametern  $n$  und  $p$ , man schreibt  $X \sim \text{Bin}(n, p)$ . Für den Erwartungswert und die Varianz von binomialverteilten Zufallsvariablen gilt

$$E(X) = E\left(\sum_{i=1}^n X_i\right) \stackrel{\text{lin.}}{=} \sum_{i=1}^n E(X_i) = n \cdot p$$

$$V(X) = V\left(\sum_{i=1}^n X_i\right) \stackrel{\text{unabh.}}{=} \sum_{i=1}^n V(X_i) = n \cdot p(1-p)$$

### 2.1.4 Hypergeometrische Verteilung

Man betrachtet eine Grundgesamtheit der Größe  $N$ , in der  $M$  Elemente eine gewünschte Eigenschaft besitzen, d.h. einen Treffer darstellen.

Eine Zufallsvariable die bei einer Stichprobe der Größe  $n$  (ohne Zurücklegen) die Anzahl Treffer zählt heißt dann *hypergeometrisch verteilt*, man schreibt  $X \sim \text{Hyp}(n, M, N - M)$ .

$$P(X = k) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$$

Sei  $A_j$  das Ereignis, dass die  $j$ te Kugel die gewünschte Eigenschaft hat. Die Zufallsvariable lässt sich also schreiben als

$$X = \sum_{j=1}^n 1_{A_j}$$

Die Wahrscheinlichkeit, für  $A_j$  ist  $P(A_j) = \frac{M}{N}$ . Damit ist der Erwartungswert einer hypergeometrisch verteilten Zufallsvariable

$$E(X) = E\left(\sum_{j=1}^n 1_{A_j}\right) = \sum_{j=1}^n E(1_{A_j}) = n \cdot \frac{M}{N}$$

## 2.2 Erweiterung auf allgemein diskrete Wahrscheinlichkeitsräume

Bisher haben wir endliche Wahrscheinlichkeitsräume betrachtet, das heißt  $0 < |\Omega| < \infty$ . Wir erlauben nun eine Erweiterung auf  $0 < |\Omega| \leq |\mathbb{N}|$ .

### Definition 2.15: Diskreter Wahrscheinlichkeitsraum

Das Paar  $(\Omega, P)$  ist ein diskreter Wahrscheinlichkeitsraum, wenn

**Nicht Negativität**  $\forall A \subseteq \Omega : P(A) \geq 0$

**Normiertheit**  $P(\Omega) = 1$

**$\sigma$ -Additivität** Für alle paarweise disjunkten Teilmengen  $A_1, A_2, \dots \subseteq \Omega$  gilt  $P(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$  gilt.

**KONSEQUENZEN AUS DER ERWEITERUNG:**

- Endliche Wahrscheinlichkeitsräume sind Sonderfälle von diskreten Wahrscheinlichkeitsräumen.
- Es gilt wie im endlichen Fall  $P(A) = \sum_{\omega \in A} P(\{\omega\})$ . Hier kann aber potentiell  $A$  unendlich viele Elemente beinhalten, allerdings handelt es sich stets um eine absolut konvergente Reihe.
- Alle Rechenregeln für Wahrscheinlichkeiten gelten weiter.
- Aber  $E(X) = \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\})$  ist eventuell eine nicht konvergente Reihe. Der Erwartungswert einer Zufallsvariable ist nur definiert, wenn  $\sum_{\omega \in \Omega} |X(\omega)| \cdot P(\{\omega\}) < \infty$ .

**BEISPIEL:** Das sogenannte *Sankt-Petersburg-Paradoxon* behandelt das Problem von nicht existierendem Erwartungswert. Es wird eine Münze geworfen bis zum ersten Vorkommen von Zahl im Wurf  $k$ . Der Grundraum ist also  $\Omega = \mathbb{N}$ . Damit ist  $P(\{k\}) = \frac{1}{2^k}$ , die Zufallsvariable  $X$  beschreibt die Auszahlung in Abhängigkeit von  $k$ ,  $X(k) = 2^k k - 1$ . Die Reihe

$$\sum_{k \in \mathbb{N}} X(k) \cdot P(\{k\}) = \sum_{k=0}^{\infty} \frac{1}{2} \rightarrow \infty$$

ist divergent - der Erwartungswert  $E(X)$  existiert nicht.

**2.2.1 Geometrische Verteilung**

Wir führen nun unabhängige Bernoulli-Experimente mit Trefferwahrscheinlichkeit  $0 < p < 1$  so oft hintereinander aus, bis ein Erfolg eintritt. Damit ist der Grundraum die Menge

$$\Omega = \left\{ \underset{\omega_1}{1}, \underset{\omega_2}{01}, \underset{\dots}{001}, 0001, \dots \right\}.$$

$\Omega$  ist abzählbar unendlich groß. Aufgrund der Unabhängigkeit der einzelnen Experimente ist die Wahrscheinlichkeit für ein Ergebnis

$$P(\{\omega_k\}) = (1-p)^{k-1} \cdot p = q^{k-1} p.$$

Die Zufallsvariable, die die Anzahl Misserfolge bis zum Erfolg zählt  $X(\omega_k) = k-1$  heißt dann *geometrisch verteilt* mit Parameter  $p$ , in Zeichen  $X \sim G(p)$ .

Es ist also

$$P(X = k) = q^k p.$$

Die Verteilungsfunktion an den Stellen  $i \in \mathbb{N}$  der geometrisch Verteilten Variable ist

$$F^X(i) = P(X \leq i) = \sum_{k=0}^i P(X = k) = p \cdot \sum_{k=0}^i q^k = p \cdot \frac{1 - q^{i+1}}{1 - q} = 1 - q^{i+1}.$$

Und analog gilt für den Erwartungswert

$$E(X) = \sum_{i=0}^{\infty} i \cdot P(X = i) = \sum_{i=0}^{\infty} i \cdot q^i p = pq \cdot \sum_{i=0}^{\infty} i \cdot q^{i-1} \stackrel{2.1}{=} p \cdot q \cdot \frac{1}{(1-q)^2} = \frac{p \cdot q}{p^2} = \frac{q}{p} = \frac{1}{p} - 1$$

Mit der Umformung

$$\sum_{i=0}^{\infty} i \cdot q^{i-1} = \sum_{i=0}^{\infty} \frac{d}{dq} q^i = \frac{d}{dq} \sum_{i=0}^{\infty} q^i = \frac{d}{dq} \frac{1}{1-q} = \frac{1}{(1-q)^2}. \quad (2.1)$$

**BEISPIEL:** Wir betrachten ein Spiel mit Sammelbildern (z.B. Fußballspieler aus Hanuta). Die Zufallsvariable  $Y_i$  beschreibt die Anzahl Bilder, die man ziehen muss um seine Sammlung von  $i$  Bildern auf  $i + 1$  zu vergrößern.

Damit ist die Anzahl an Misserfolgen vor dem richtigen Bild geometrisch verteilt

$$(Y_i - 1) \sim G\left(\frac{n-i}{n}\right) \Rightarrow E(Y_i) = E(Y_i - 1) + 1 = \frac{1}{\frac{n-i}{n}} = \frac{n}{n-i}$$

Insgesamt beschreibt die Zufallsvariable  $Y$  die benötigten Bilder für eine Sammlung der Größe  $n$

$$Y = \sum_{i=0}^{n-1} Y_i \Rightarrow E(Y) = \sum_{i=0}^{n-1} E(Y_i) = \sum_{i=0}^{n-1} \frac{n}{n-i} = n \cdot \sum_{i=1}^n \frac{1}{i} \approx n \cdot \ln(n)$$

Für maximal 63 mögliche Bilder ist  $E(Y) = 297.88$  und die Näherung  $63 \ln(63) = 261.01$ .

### Satz 2.16: Gedächtnislosigkeit der geom. Verteilung

Für eine geometrisch verteilte Zufallsvariable  $X \sim G(\lambda)$  und  $x, y \in \mathbb{N}_0$  gilt

$$P(X \geq x + y \mid X \geq x) = P(X \geq y)$$

das heißt,  $X$  ist gedächtnislos.

**BEWEIS:**

$$\begin{aligned} P(X \geq x + y \mid X \geq x) &= \frac{P(X \geq x + y) \cap P(X \geq x)}{P(X \geq x)} \\ &= \frac{P(X \geq x + y)}{P(X \geq x)} \\ &= \frac{1 - P(X \leq x + y - 1)}{1 - P(X \leq x - 1)} \\ &= \frac{1 - F^X(x + y - 1)}{1 - F^X(x - 1)} \\ &= \frac{q^{x+y}}{q^x} \\ &= q^y = P(X \geq y) \end{aligned}$$

**Bedeutung:** Diese Aussage widerspricht der Intuition, dass zum Beispiel beim Würfeln die Wahrscheinlichkeit eine 6 zu werfen größer werden muss, wenn schon lange keine geworfen wurde.

## 2.2.2 Poisson-Verteilung

Eine Zufallsvariable  $X$ , die Ereignisse in einem Zeitintervall zählt, wobei im Mittel  $\lambda$  Ereignisse auftreten heißt *Poisson-verteilt* mit Parameter  $\lambda$ , in Zeichen  $X \sim \text{Po}(\lambda)$ .

**HERLEITUNG:** Die Poisson-Verteilung entsteht als Grenzwert für  $n \rightarrow \infty$ , wenn man das betrachtete Zeitintervall in  $n$  Teile der Länge  $1/n$  aufteilt, so dass in jedem Teilintervall genau 1 oder 0 Ereignisse auftreten.

Die Anzahl der Ereignisse  $X_n$  im gesamten Intervall ist damit  $\text{Bin}(n, \frac{\lambda}{n})$ -verteilt. (Hieraus folgt,  $E(X_n) = \lambda$ .)

Für  $X_n$  gilt dann nach der Binomialverteilung

$$\begin{aligned}P(X_n = k) &= \binom{n}{k} \cdot p_n^k \cdot (1 - p_n)^{n-k} \\&= \frac{n^{\underline{k}}}{k!} \cdot \frac{(n \cdot p_n)^k}{n^k} \cdot \left(1 - \frac{n \cdot p_n}{n}\right)^{-k} \cdot \left(1 - \frac{n \cdot p_n}{n}\right)^n \\&= \frac{n^{\underline{k}}}{k!} \cdot \frac{\lambda^k}{n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^{-k} \cdot \left(1 - \frac{\lambda}{n}\right)^n \\&= \frac{\lambda^k}{k!} \cdot \frac{n^{\underline{k}}}{n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^{-k} \cdot \left(1 - \frac{\lambda}{n}\right)^n\end{aligned}$$

für  $k \rightarrow \infty$  folgt

$$\begin{aligned}P(X_n = k) &= \frac{\lambda^k}{k!} \cdot \underbrace{\frac{n^{\underline{k}}}{n^k}}_{\rightarrow 1} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\rightarrow 1} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\rightarrow e^{-\lambda}} \\&= \frac{\lambda^k}{k!} \cdot e^{-\lambda}\end{aligned}$$

Die Verteilungsfunktion konvergiert damit gegen 1, denn

$$\begin{aligned}F^X(k) &= \sum_{i=1}^k \frac{\lambda^i}{i!} \cdot e^{-\lambda} \\ \lim_{k \rightarrow \infty} F^X(k) &= \sum_{i=1}^{\infty} \frac{\lambda^i}{i!} \cdot e^{-\lambda} = e^{-\lambda} \cdot \sum_{i=1}^{\infty} \frac{\lambda^i}{i!} = \frac{e^{\lambda}}{e^{\lambda}} = 1.\end{aligned}$$

**FOLGERUNG:**

- Für die Wahrscheinlichkeiten einer Poissonverteilten Zufallsvariablen gilt

$$P(X = i) = \frac{\lambda^i}{i!} \cdot e^{-\lambda} \quad i \in \mathbb{N}_0$$

- Die Varianz ist, wie der Erwartungswert  $V(X) = E(X) = \lambda$ .
- Für zwei unabhängige Zufallsvariablen  $X \sim \text{Po}(\lambda)$  und  $Y \sim \text{Po}(\mu)$  ist

$$(X + Y) \sim \text{Po}(\lambda + \mu)$$

## 2.3 Grenzwertsätze

# LITERATURVERZEICHNIS

- [1] L. FAHRMEIR, *Statistik: Der Weg zur Datenanalyse*, Springer-Lehrbuch : SpringerLink : Bücher, Springer Spektrum, Berlin, Heidelberg, 8. Aufl. 2016 ed., 2016.
- [2] N. HENZE, *Stochastik für Einsteiger: Eine Einführung in die faszinierende Welt des Zufalls*, SpringerLink : Bücher, Springer Spektrum, Wiesbaden, 10., überarb. Aufl. 2013 ed., 2013.
- [3] U. KRENGEL, *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, Vieweg Studium, Aufbaukurs Mathematik : SpringerLink : Bücher, Vieweg+Teubner Verlag, Wiesbaden, 8., erweiterte Auflage ed., 2005.